DOCUMENT RESUME

ED 375 816 IR 016 872

AUTHOR Gasser, P.
TITLE Audio Card Systems. Technical Information Bulletin No. 13.
INSTITUTION International Council for Educational Media, The Hague (Netherlands).
PUB DATE Jun 94
NOTE 18p.; This document is an abstract in English of the dossier "Les cartes vocales" (June 1993) written by the "Direction de l'Ingenierie educative" of the "Centre National de Documentation Pedagogique" of France.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Acoustics; *Computer Uses in Education; Elementary Secondary Education; Foreign Countries; Futures (of Society); Sound Effects; *Speech Synthesizers; Technological Advancement; *Word Recognition
IDENTIFIERS France; International Council for Educational Media

ABSTRACT

This examination of audio card systems for computers begins by identifying the three information processing systems for sound: sound digitizing, synthesis of text, and word recognition. Specific pedagogical applications of digitized sound are then briefly discussed. The remainder of the document focuses on specifications for the working of vocal cards. This is divided into the three techniques mentioned above. The first section is concerned with digitizing sound and covers topics such as sampling, quantification, storage and capacity, compression, and filters. The second section concentrates on speech recognition by computers, including obstacles to vocal recognition, speaker systems, discontinuous speech or enforced continuous speech, recognized vocabulary, error rate and rejects, and restrictions of use. Discussion of speech synthesis starting from texts in the third section covers its current state of development and possible future educational applications. (JLB)

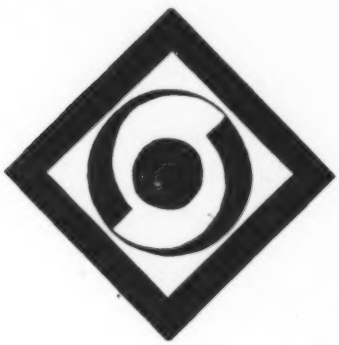
****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

INTERNATIONAL COUNCIL FOR EDUCATIONAL MEDIA
CONSEIL INTERNATIONAL DES MOYENS D'ENSEIGNEMENT

ED 375 816

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ICEM TECHNICAL INFORMATION BULLETIN



AUDIO CARD SYSTEMS

ICEM Technical Information
Bulletin N° 13 - June 1994

Prepared by P. Gasser

Direction de l'Ingénierie éducative
Centre National de Documentation
Pédagogique - France

Issued by

The Subcommittee on
Equipment and Management

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard A. Cornell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

BEST COPY AVAILABLE

This document is an abstract in English of the dossier "Les cartes vocales" - (Juin 1993) written by the "Direction de l'Ingénierie éducative" of the "Centre National de Documentation Pédagogique" of France. ICEM members are licensed to make, in whole or in part, their own translations and copies provided that due acknowledgement is given to CNDP and ICEM.

Contents

		Page
1	**Three information processing systems of sound**	**1**
2	**A new tool for education**	**2**
2.1	Pedagogical applications : panorama	2
2.2	Vocal synthesis : experiences taken from life	2
2.3	The language laboratory is being computerized	3
3	**Working principles**	**3**
3.1	From analogy to digitizing	3
3.1.1	Sampling	5
3.1.2	Quantification	6
3.1.3	Quantification noise	6
3.1.4	Storage and capacity	7
3.1.5	The compression	7
3.1.6	Restitution	8
3.1.7	The filters	8
3.2	Speech recognition	8
3.2.1	Obstacles to by-pass	9
3.2.2	Monospeaker or multispeaker systems	9
3.2.3	Discontinuous speech or enforced continuous speech	10
3.2.4	The recognized vocabulary	11
3.2.5	Error rate and rejects	11
3.2.6	Restrictions of use	11
3.3	Speech synthesis starting from texts	12
3.3.1	A general survey of the working of vocal synthesis	13
3.3.2	Diphones	13

AUDIO CARD SYSTEMS

1 Three information processing systems of sound

Go into an information fair. A well-known actor's voice speaks to you, you answer it and that is the beginning of a dialogue with a machine. Furthermore, you dictate your text and you receive it all printed out. More than that, you put it on a scanner; a feminine voice will disperse it all round... Enter in the world of the all-digital and discover how the computers'sonorous capacities are used. As a guideline, we present the three information processing systems applied :

- **Sound digitizing**

  The computer is above all able to reproduce sounds of natural origin. The sonorous message, entirely pronounced by a human speaker or created by musical instruments, is recorded on an analog carrier or caught by a microphone. An information system (sampler) makes it possible to convert and to encode it as digital data captured on analog carriers (disks, hard disks, CD-ROM...) or on audio-digital carriers (audio compact disks). Those data can be compressed and finally be reproduced in analog form by loudspeakers, amplifiers, etc. In this case, we talk about digitizing or sound reproduction techniques.

- **The synthesis of texts**

  The computer can also produce synthetic sounds, wholly or partly created by a computer. It concerns for instance electronic music coming from electronic instruments, which communicate with each other and with the computer by way of the MIDI standard. In this case it concerns above all speech synthesis starting from texts where the vocal sign is completely created starting from a written text, with the help from a dictionary of minimal acoustic signals recorded and digitized in advance. A set of word production rules adds intonation, prosody, connections, etc. Listening to the message, it is perfectly audible and intelligible, but it has still an artificial character which is rather obvious. Techniques to synthesize texts have much developed during the last few years; although they still offer enough material for further research.

- **Word recognition**

  The computer finally seems to "understand" and to "obey" the voice. After having digitized the message, the computer has to compare it with an acoustic vocabulary of reference, from which resemblances are extracted and from which a certain understanding of what has been told is deduced. The word recognition techniques, which are in full development, are finding an outlet.

2 A new tool for education

As long as computers could only handle digital or textual data, it was difficult to allow for the sonorous and vocal dimension in educational computer science.

Nevertheless, first experiments combining information tools and audio-visual carriers saw the light at that time. They were based on an instrument operating an audiotape recorder, slides, a videotape recorder or a videodisc, controlled by a computer. They made it possible to benefit from specific ways of data processing, the setting up of a permanent interactivity between machine and user, and complementary contribution by audio-visual techniques.

Nowadays, the arithmetical power of informatics is used for a radical change of the working environment, which is even from the beginning interactif, convivial and multimedia. The audio-visual elements are gradually integrated in the computer, which enables it to produce and reproduce all sorts of sounds, but also fixed images, cartoons and sequences of video images. The computer screen has lost its austere character and resembles more and more a television screen, but a television which can be programmed and explored by the user himself, who is looking for information and ways of processing. The introduction of the digital sound, which is now controlled and attainable, creates that modification. The pedagogical applications, mentioned hereafter, are new from a technological point of view, nevertheless, they are based on a firm reality : the role of oral communication as support in the transfer of knowledge.

2.1 Pedagogical applications : panorama

The use of sound in computers finds its first application in modern language teaching. But it involves also other fields, from literature to physics, from music teaching to biology. In general, it is a more natural way of communication between man and machine, communication which will be supported by this evolution.

2.2 Vocal synthesis : experiences taken from life

At his admission into hospital, Samson, six years old, suffers from an autistic psychosis. However, thanks to the pedagogical team surrounding him ; he awakes bit by bit from his lethargy. He becomes interested in the voice and, at nine years old, in the characters on the computer keyboard.

The educator, benefiting from the vocal synthesis, asks the child to write down one of his favourite nursery-rhymes, while dictating the characters one after the other. Afterwards, he shows the child how to make the computer read the text. Samson is amazed, he repeats it several times in succession, his eyes glued to the computer !

Then he would volunteer to work with other children, would form his first name on the keyboard (which he had never written down in the past) and would continue to learn writing without the computer.

2.3 The language laboratory is being computerized

Thanks to its new sonorous capacities, the computer is able to perform all the recorder's functions. One could think that it will gradually replace the classic language laboratories. However, it is important to understand that the computer enables the development of different activities and hence, offers a complementary tool for language teaching.

3 Working principles

Some people may believe it needless to understand the working of vocal cards. We are used to use new products without necessarily knowing the technologies on which they are based. Nevertheless, it seems essential to offer some specifications about the mechanisms applied, while making a difference between digitizing, recognition and synthesis techniques.

The digitized sound is part of everyday life, in particular with the CD audio. Whereas digitizing a text can be reduced to the encodation of the successive characters, digitizing analog signals - sound and image - is much more complex. Here, we have especially insisted on the points which the user might consider parameters of sound digitizing : by controlling those elements, he can better grasp the technical working plans added to the cards, and choose, according to his needs, how the different sonorous messages can be optimally processed : vocal annotations, words, sentences in a foreign language, music.

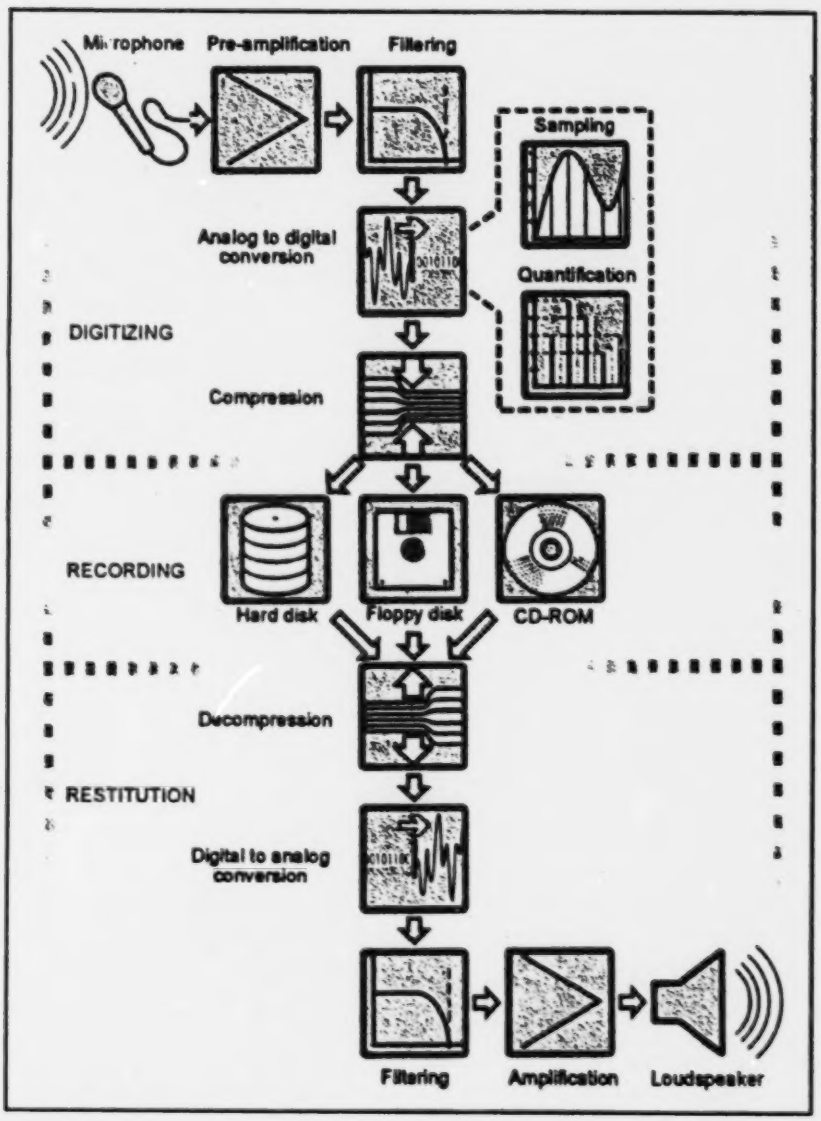
Vocal synthesis and recognition techniques are less accepted and offer still many prospects for the future. We confined ourselves to the general principles, without going into technical details. Whereas school institutions hardly ever have applications of these techniques at their disposal, it is important to describe the extent of applications used in daily life or in a working environment. The numerous examples should lead to a better understanding of these possibilities, which from now on enable us to think about the use of certain applications in education.

3.1 From analogy to digitizing

The recording of a sound is based on the continuous variations of the electric signal produced by the vibrations of the microphone's diaphragm. In the case of a recorder, the vibrations are transformed in variations produced by the magnetizing of a magnetic carrier : it concerns analog recording.

For digital recording, the electric signal is encoded in a series of digital values. The transformation requires first of all the cutting of the signal into times with a regular time interval, i.e. sampling, and then a quantification of each sample, obtained by comparing with reference values which were defined in advance. To those two operations, which occur practically simultaneous, we should add the filtering of the digital signal and its recording on an information carrier.

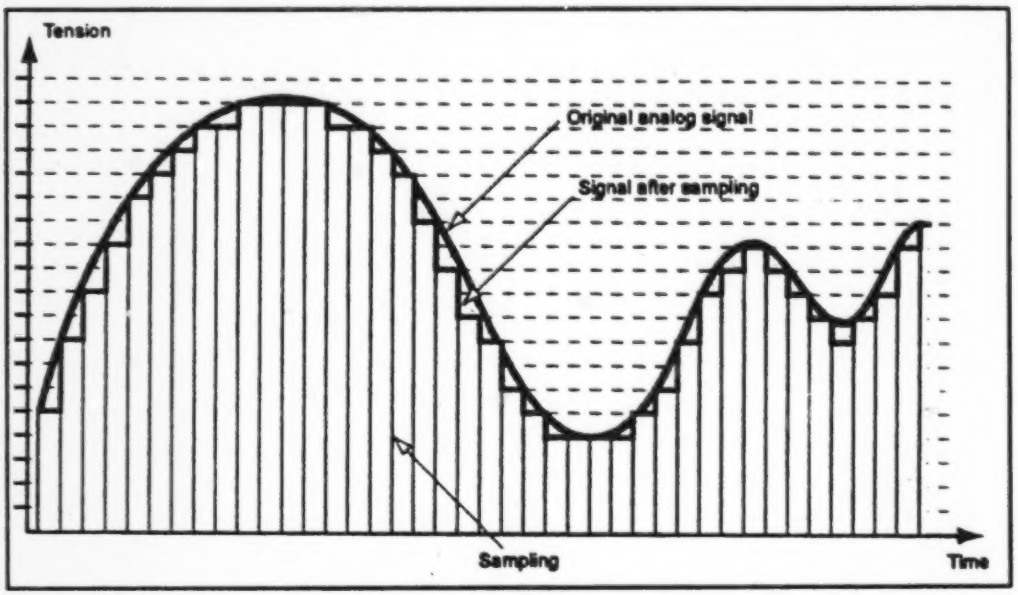
The transformation of the digital sound into an audible signal requires a reproduction operation, which is opposite from the previous one. The digital data are converted into variations of an electric signal which are amplified and transmitted to a loudspeaker.

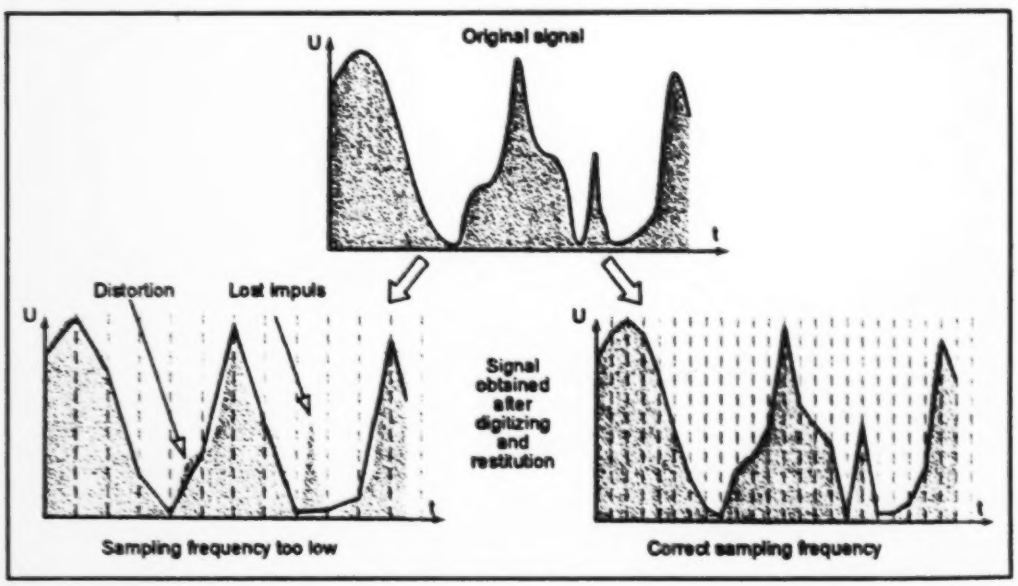


3.1.1 Sampling

The audio signal which has to be digitized is represented by a continuous curve and shows no real regularity. In a first stage, the analog signal is temporary to be divided up in fine slices, so that the amplitude reached in each slice can be measured : this process is called sampling.

The frequency of sampling or the number of samples per second (measured in hertz) should be in proportion to the maximal frequencies which can be reached by the audio signal. If the frequency is too weak, the signal's quick variations at the beginning will not be recorded and hence be lost forever : the restituted signal will be distorted. Shannon proved that to encode a certain signal, the sampling frequency should be at least equal to twice the maximal frequency of the signal.





BEST COPY AVAILABLE

3.1.2 Quantification

During this operation, the amplitude of every single sample in the first stage is evaluated in order to convert them into digital values. As for this conversion, we have to define a reference interval in order to cover each possible value. This interval is subdivided in a series of N bearings, thus forming a gradation (regular in case of uniform or linear quantification). The amplitude of every sample is represented by a whole number lower than N, in comparison with the gradation established. Note that the amplitude of an analog signal varies in a continuous way and takes an infinite number of values (between two extremes), while after being digitized the signal will only be represented by a limited number N of values, which are called quantums.

The digital values obtained this way are encoded by means of a fixed number n of bits (4, 8 or 16 in general), just like all other memorized numbers in computers. The number N of different levels which reproduces the sound's amplitude depends on the choice of n ($N = 2^n$). A system encoding on 8 bits offers 2 to the power of 8, i.e. 256 quantums ; when it encodes on 12 bits, we obtain 65.536 quantums.

3.1.3 Quantification noise

When digitizing, one single value may correspond to different values approaching the analog signal, which causes quantification errors; producing a random noise, called quantification noise. The lower the number of bits per sample, the greater the risk of errors.

This error is merely caused by the conversion in number of every sample. It is not depending on the sampling frequency, but only on the number of bits chosen to encode the digital values, i.e. in fact the resolution power of the analog digital converter.

The quantification error cannot be corrected (the restituted audio signal will be degraded), but can be calculated ; the maximum percentage is one unit on the total number of quantums, (e.g. 1/256, which is 0,39% for a system sampling on 8 bits). The quantification noise is evaluated by the ratio signal - noise, measured in decibel (db) of which the theoretical value is equal to 20 log N.

As for the cards sampling on 8 bits, the theoretical value of this ratio amounts only to 48 dB. For a system encoding on 16 bits, this value improves when reaching 96 dB. To this noise characteristic for quantification, should be added deteriorations inherent in circuits treating analog audio signals.

The relation signal-noise is generally measured by another, more complex unit, the dBA, which takes the curve of the ear sensibility into account (e.g. language laboratories require a ratio higher than 56 dBA).

3.1.4 Storage and capacity

At the end of the quantification, the signal is encoded in digital form : it can be stored in the memories of the computer (disk, diskette...), or transmitted to other peripheral equipment. The capacity can be represented by the ratio D = F.n, in which D is the capacity in bits per second, F the sampling frequency and n the number of bits used to encode each sample. It corresponds also to the volume of 1 second digitized sound.

The described digitizing technique is called Pulse Code Modulation. The capacity is enforced by the system's characteristics, i.e. the number of bits used for every sample and the sampling frequency. To increase the quality of the signal, at least one of these parameters should be increased ; conversely the capacity is always at the expense of the quality of the signal.

3.1.5 The compression

We seize the importance of research concerning compression systems, in order to reduce the signal's capacity while minimizing quality loss of the initial signal. These techniques have many advantages : as for the digitizing/restitution cards, the files' volume stored in the computer memories and a quicker access ; as for the communication systems (local or public networks), improvement of the transit conditions.

The most important compression technique uses the redundancies of the audio signal and encodes only the differences in amplitude between two successive samples. Except for extreme variations of the signal, this difference gives indeed a less important number than the total amplitude and can hence be encoded on less bits. This method is called DPCM (Differential Pulse Code Modulation).

In order to compress the information still more, this method is improved by introducing a process of extrapolation of the signal's variation speed, in order to adapt the number of bits necessary for encoding to the amplitude of that difference. This method is called ADPCM (Adaptive Differential Pulse Code Modulation).

Other compression techniques are based on the fact that it is not necessary to encode anything that cannot be heard by the ear, because it is below the threshold of auditif perception ; or because one sound is masked by another one, a weak sound for instance cannot be heard when it is dominated by a strong noise.

3.1.6 Restitution

The digital signal, possibly restituted after decompression, is reconstructed in analog form after a double operation :

- a digital/analog conversion transcribes the succession of digital codes in an electric signal;
- an output filter (low-passfilter) reconstitutes the original signal by eliminating any constituent of parasitic frequencies exceeding half of the sampling frequency; thanks to this filter, the signal will loose its "angular" appearance and become more rounded, hence the term smoothing filter.

3.1.7 The filters

As explained before, the mathematical models described for sound digitizing show that a system can only digitize correctly sounds of which the frequency is lower than the limit value. This limit is about half of the sampling frequency. Two additional phenomena can be noticed :

- Digitizing sounds with frequencies exceeding this limit value causes parasitic frequencies, which may be audible in the restitution. The sound should thus be filtered before being digitized in order to eliminate frequencies causing this problem. Therefore we use a low-passfilter at the entry.
- Digitizing sounds with a frequency approaching this limit value may cause parasitic frequencies exceeding the limit, but audible when restituting the sound. The restituted sound will be filtered at the exit, also by means of a low-passfilter.

3.2 Speech recognition

"The computer can listen to you, it obeys your voice ; it notes down while you are dictating..; you can do without a keyboard, a mouse... even without a secretary !" According to the press, a revolution is imminent and indeed, the efforts by different research centres seem to produce a significant progress in speech recognition techniques. The systems are developing speech recognition under circumstances that are close to everyday reality. The speaker faces less restraints concerning the recognition of his word. And what is more, the systems are now able to recognize a greater number of different speakers, while reducing the error rate. English is not any longer the only language involved and products using the French language, which is repudiated to be very difficult, are coming onto the market.

In a first stage, the use of vocal control was restricted to the world of the disabled or to elementary operations. Nowadays, it also concerns the general public, as shows the recent progress in telephony with the growing development of services accessible by means of vocal servers, and the use of vocal controlling modules on sonorous cards.

Few applications concern education. Although some prospection regarding pedagogical research and experimentation may be useful. A realistic analysis of present progress in these techniques and their possible applications is necessary in order to take part consciously in their applications.

3.2.1 Obstacles to by-pass

Automatic speech recognition presupposes two essential and complementary functions. On the one hand, the machine should track speech segments (words, sounds) by identifying and comparing the data obtained after digitizing the pronounced chain, with acoustic elements of a reference vocabulary, also digitized. On the other hand, it should "comprehend" the pronounced message, it should attribute at least one meaning to the identified form. With this last point we return to the more general issue of natural language comprehension, a subject treated by Artificial Intelligence techniques, and still offering a vast area for research.

Automatic speech recognition runs up against the problem of identifying an acoustic form without relying at the same time on the comprehension of the message that has to be decoded, unlike the functioning of the brain in which those two functions are not separated...

That is why variations of the voice (tiredness, irritation, emotion...), as well as differences in accent, in intonation or in speed of speech, make speech recognition even more complicate. Prosody, intonation, the use of breaks, links and other forms of coarticulation do not lead to one single model of segmentation. What is more, in order to compare the pronounced message with the acoustic elements of the reference dictionary, it has to be divided up into isolated elements.

A microphone does not offer the selective possibilities of the system "brain - ear" and the background noise is mixed up with the spoken chain. The acoustic signal is also damaged by the digitizing system, because of its traffic on electric circuits and telephone lines.

We still do not control all these difficulties, that is why vocal recognition offers an enormous field for research. The modern systems by-pass these problems by using makeshifts, hereinafter presented.

3.2.2 Monospeaker or multispeaker systems

Every vocal recognition system requires a learning period during which the system grows accustomed to the speaker's voice; but the system can be different according to the preconceived application : some systems adapt themselves to every single speaker, others respond immediately to whatever speaker (this concerns telephonic applications).

In the case of a monospeaker application, the system submits each speaker to a learning period of its voice. The application is personalized by introducing vocal reference files characteristic for each user. The process used differs according to the system; for some systems the lexicon has to be repeated several times, word after word, others control the learning process by ending it as soon as the recognition rate is satisfactory, still others try to make it transparent by adapting themselves progressively to the speaker, thus improving their achievement during the first hours or days of use.

A multispeaker system is supposed to recognize immediately whatever speaker, without preparatory teaching period of every voice, by which is meant a recognition "statically independed from the speaker". This target is achieved when the recognition rate is about 80% to 90%, whatever speaker. However, a preparatory learning period is necessary to equip the multispeaker system with a sufficient number of representative voice samples, or with a statistic model representing the whole of acoustic forms found in the voice sampling of hundreds of people. We talk about average learning from a set of speakers.

3.2.3 Discontinuous speech or enforced continuous speech

Modern techniques do not allow for the recognition of natural speech. The difficulties met are by-passed by imposing some restrictions on the speaker, which makes recognition easier.

By imposing for instance an elocution in which words are isolated, the system charges the speaker with the segmentation of the sonorous message. The speaker should wait for the system's reaction after every pronounced word (technique of isolated words, used for systems under vocal control or for vocal data capture), or compel himself to pronounce each word separately, by introducing a break between the words and by avoiding links (vocal dictation, called discontinuous speech).

The technique of enforced continuous speech asks the user to respect a fixed syntax and a limited vocabulary.

The speaker can pronounce complete sentences using a normal rhythm, but corresponding to a precise application. For instance, among the sentences used during a meteorological speech, we can determine the recognition of those respecting the syntax : "What wheather + TO BE + WHEN"; "TO BE" corresponds to the corpus "will be, would be, was...", "WHEN" corresponds to "today, tomorrow, this week, this summer..." or "the + NUMBER + MONTHS", by combining in the same way the definition of NUMBER, MONTHS, etc.

One of the aims pursued by recent research is to reduce these restrictions.

3.2.4 The recognized vocabulary

The extent of the recognized vocabulary is a parameter for every recognition system. For monospeaker use and under good conditions (without any noise, in a room), some systems can recognize at a certain moment several thousands of isolated words or several hundreds in enforced speech. On the other hand, in telephonic applications, for multispeaker use, in bad circumstances (noise, parasites...) the achievements are only of the order of a dozen words. That does not mean that there aren't any limited applications which recognize no more than ten words. Ramification of dialogues or enforced continuous speech are means to limit at a certain moment the subset of the words to be recognized and to anticipate the next one to handle with. The overall extent of the treated vocabulary has considerably increased this way.

3.2.5 Error rate and rejects

When the system takes one word for another, we talk about an error. When it refuses a word which is too remote from its vocabulary, it concerns a reject. A system is more efficient when it prefers to do nothing and to wait, instead of doing a wrong action. Certain systems add to the recognition a coefficient of confidence. The person who works out the application has all decision liberty according to the thresholds reached.

3.2.6 Restrictions of use

The use of speech recognition systems is still limited for the speaker, his speed of speech should be constant, he should talk clearly and have a good articulation, he should avoid changes of pitch and his prosody and diction are to be slightly unnatural, he should learn to separate his words, etc. An application which is completely vocal is only possible for a relatively short term (e.g. telephonic consultation), but several hours of dialogue may be intolerable for some speakers.

A recognition system can not be obtained just like that. An adaptation is needed according to the system. We have to define the application and the interfaces which will allow to use the results obtained by recognition in the appropriate working environment (text editor, post administrator, dialogue control...). The lexicon necessary for the application, and sometimes the syntax model should be studied. According to the case, we have to create a vocabulary and the corresponding acoustic dictionary or the existent data proper to the application. Next, it will be useful to measure the recognition rate, the reject rate (words non-recognized) and the error rate (words mixed up) on a sample of representative persons and in the user's context (surrounding noise, real speakers...).

For the moment, the best results are achieved for discontinuous speech, monospeaker use and in an environment which is isolated from the surrounding noise. Soft- and hardware are still relatively expensive. Though, the recognition techniques are crossing a threshold which enables the laboratories to develop some interesting products.

3.3 Speech synthesis starting from texts

"La lectrice 2 !" In the next cinematographic adaptation of his book, Raymond Jean will prefer a computer to Miou-Miou ! Its voice, perfectly comprehensible, may become slightly monochord and tiring in the end... But still : to be able to listen to a text, which hasn't been read nor recorded, is still marvelling... for naive people, at least.

The texts to be treated by a speech synthesis system can be captured by text processing software, or digitized starting from a printed document, with the help of a scanner or recognition software for characters. Then, a program analyses the chain of written characters and transforms it, by means of complex processes which are very briefly described, first into a digital signal and then in an analog one.

This technique is still very expensive. The applications are rare and concern mainly the partially sighted.

As far as education is concerned, there are some interesting experiments done concerning the use of vocal synthesis in the learning and the perfectioning of the written language at nursery school, at primary school and for the instruction of illiterate adults.

For the moment, and according to specialists, the quality obtained does not allow the use in language education. Though, it seems very important for teachers to follow the evolutions in this area very closely and to widen the field of research, while taking into account the achieved results. For instance, the synthesis reading of texts, in the mother language, of line helps, of instructions, of texts, of screens... may make the use of the computer easier for pupils with problems. After some time, the technique may turn out to be less expensive than the use of digitizing and restitution techniques of sound.

The joint evolution of synthesis and speech recognition techniques, of which the quality is improving, enables us to anticipate mutations of the interfaces "man/machine", as we will be using computers which are able to react to vocal orders, to capture texts dictated orally and to restitute them then artificially in an audible form.

3.3.1 A general survey of the working of vocal synthesis

The production of a vocal message is achieved through several methods, briefly described hereafter :

- **The synthesis by words** : a word dictionary is recorded, and then digitized. The computer puts these words together to form sentences preserving a modelization of transitions and a generation of the prosody of the pronounced sentence. The result is acceptable but not very natural, because of the last two points which are still not completely under control. Moreover, the setting-up of a dictionary is difficult and requires a lot of memory space.
- **The synthesis by minimal phonetic units** (phonemes or diphones, according to the method) : in a first stage, the text has to submit a linguistic treatment. A syntactic and grammatical analysis, accompanied by the use of a lexicon, transcribes it first into a chain of phonetic characters. Then, a prosody is generated. Thus, we obtain a "phonetic-prosodic" representation of the text. In a second stage, called signal processing, this representation is transformed into a digitized speech signal and then restituted in analog form.

3.3.2 Diphones

The CNET develops a synthesis in French in which the combined acoustic elements are diphones. Diphones are preferable to phonemes as the coarticulation between two phonemes is very difficult.

The French language has about 1200 diphones with an average length of 100 ms. The Psola cards, presented by Elan Informatique and Xcom are products of this research. Versions in other languages (English, Spanish) are being developed.

The synthesis by rules and formants or by linear prediction : it is based on a simulation system of the vocal canal thanks to the control of a set of formants (parameters of the acoustic filter) by rules.

At this moment however, the synthesis by diphones is mostly used : the diction obtained is less natural but more comprehensible.

The products of synthesis are specific for a given language. Some of them are used for English applications, like for instance the software provided by the Pro Audio Spectrum or Sound Blaster cards. Some of them are adapted to different languages.

Some specialised retailers present a work station composed of a scanner, software for character analysis, a card for speech synthesis. A scanned document (newspaper, book, etc.) is read by the computer.

There is also software which can control by means of text synthesis, the reading of the screen under DPS or in other applications, such as Word, Le Robert électronique, etc. The consultation of an encyclopedia becomes more accessible to visually handicapped people. The use of text processing systems like Word can become totally vocal, if a vocal command is added.